Row-Ordering Schemes for Sparse Givens Transformations. II. Implicit Graph Model*

Alan George
Department of Computer Science
University of Waterloo
Waterloo, Ontario, Canada N2L 3G1

Joseph Liu
Department of Computer Science
York University
Downsview, Ontario, Canada M3J 1P3

and

Esmond Ng
Department of Computer Science
University of Waterloo
Waterloo, Ontario, Canada N2L 3G1

Submitted by Richard A. Brualdi

ABSTRACT

A new graph model is presented to study the row annihilation and row ordering problems in the QR decomposition of sparse matrices using Givens rotations. The graph-theoretic results obtained can be used to derive good row orderings for certain column orderings, such as width-1 and width-2 nested dissection orderings. This model is different from the bipartite-graph model introduced in [6]. We refer to the new model as implicit because the rows are not represented explicitly by nodes, in contrast

*The research was sponsored by the Canadian Natural Sciences and Engineering Research Council under Grants A5509 and A8111, and was also sponsored by the Applied Mathematical Sciences Research Program, Office of Energy Research, U.S. Department of Energy under contract DE-AC05-84OR21400 with the Martin Marietta Energy Systems Inc., and by the U.S. Air Force Office of Scientific Research under contract AFOSR-ISSA-84-00056.

52 Vanderbilt Ave., New York, NY 10017 0024-3795/86/$3.50

to the bipartite-graph model, where the rows are represented by nodes in a bipartite graph.

---

1. INTRODUCTION

In [6], we presented a model which uses bipartite graphs to study the problem of ordering the rows in the QR decomposition of a sparse $m \times n$ $(m \geqslant n)$ matrix A using Givens transformations. Based on this bipartite-graph model, we showed that good row orderings can be obtained for both the conventional (width-1) nested-dissection and the so-called width-2 nested-dissection column orderings. The model is *explicit* in the sense that the rows of the original matrix A are represented by nodes in a bipartite graph. Furthermore, the bipartite-graph model describes the reduction of the matrix to upper trapezoidal form in a column-by-column manner [6].

In this paper, we introduce another graph model to study the row-ordering problem. This new model is *implicit*, since the original matrix A is not explicitly represented. The model is obtained by considering the matrix A^TA and is based on a graph model for the Cholesky decomposition of sparse symmetric positive definite matrices. The implicit graph model describes the sequence of upper triangular matrices obtained in the QR decomposition of A using Givens transformations, when A is reduced to upper trapezoidal form in a row-by-row manner. The main advantages of the implicit scheme are that it models the way in which the sparse matrix A is actually reduced to upper trapezoidal form and it provides a mechanism to study the row-ordering problem in a natural manner. Using this model, we can also show that good row orderings can be obtained for both width-1 and width-2 nested-dissection column orderings, just as was shown in [6].

An outline of this paper is as follows. In Section 2, we review the use of Givens transformations in the QR decomposition of sparse matrices and provide some basic results. The implicit graph model is then introduced in Section 3, along with some graph-theoretic results. In Section 4, we examine the various row-ordering strategies for width-1 and width-2 nested-dissection column orderings. Finally, some concluding remarks are provided in Section 5.

Some of the results presented in Sections 2 and 3 appear in [8] using matrix notation.

We assume that the reader is familiar with the basic graph-theoretic terminology used in the analysis of sparse matrix computations, such as symmetric graphs of symmetric matrices, adjacent sets, separators, and the notion of reachability [5].

2. SPARSE QR DECOMPOSITION USING GIVENS TRANSFORMATIONS

Let x and y be two sparse row vectors, and suppose x_i and y_i are both nonzero. Then one can annihilate y_i by constructing a Givens transformation (or rotation) using x_i and y_i, and applying this transformation to both x and y. Denote the *transformed* vectors by $\bar{x}$ and $\bar{y}$. Assuming exact cancellation does not occur, the *structure* of each of $\bar{x}$ and $\bar{y}$ is the union of the structures of x and y (except that $\bar{y}_i$ is now zero). Following [8], x_i and x are referred to as the *pivot element* and *pivot row* respectively. Furthermore, we define the *cost* of such an operation to be the number of nonzeros in the transformed pivot row $\bar{x}$. Note that the actual number of multiplicative operations required in such a transformation is roughly a multiple of the cost defined above.

Consider an $m \times n$ sparse matrix A, with $m \geqslant n$. We denote the kth row of A by a^k and its elements by a_j^k, $1 \leqslant j \leqslant n$. Assume A has full column rank, and denote its QR decomposition by

$$A = Q\begin{pmatrix} R \\ O \end{pmatrix},$$

where Q is an $m \times m$ orthogonal matrix and R is an $n \times n$ upper triangular matrix. One way of obtaining such a decomposition is to apply Givens transformations to the rows of A [1,3]. More precisely, we generate a sequence of $n \times n$ upper triangular matrices

$$\{R^0, R^1, R^2, \ldots, R^m\},$$

where $R^0 = O$ is the $n \times n$ zero matrix, and for $k = 1, 2, \ldots, m$, R^k is obtained from R^{k-1} by *rotating in* a^k using Givens transformations. That is, the diagonal elements and the corresponding rows of R^{k-1} are used as pivot elements and pivot rows to annihilate the nonzeros in a^k. Note that this may introduce *fill-in* in both R^{k-1} and a^k, even though the fill-in in a^k will eventually be annihilated. When all the nonzeros, including any fill-in, in a^k are annihilated, a^k is said to be *annihilated*. Since the pivot elements are chosen from the diagonal of R^{k-1}, some of them may be zero. However, it is clear from the annihilation process that if any diagonal element of R^{k-1} is zero, then the corresponding row must be entirely zero. Thus, for zero pivot element, the effect of rotating in a^k is to transfer the entire row a^k into row k of R^{k-1}. Note that $R = R^m$.

An example illustrating the transformation process is given in Figure 1. Nonzero elements of R^{k-1} and a^k are denoted by $\times$, nonzeros (fill-in)

$$R^{k-1} = \begin{pmatrix} \times & \times & & & & & \\ & \times & & & \times & & \\ & & \times & & & & \\ & & & \times & & & \times \\ & & & & \times & & \\ & & & & & \times & \\ & & & & & & \times \end{pmatrix} \qquad R^{k} = \begin{pmatrix} \otimes & \otimes & \oplus & & & & \\ & \otimes & \oplus & & \otimes & & \\ & & \otimes & & \oplus & & \\ & & & \times & & & \times \\ & & & & \otimes & & \\ & & & & & \times & \\ & & & & & & \times \end{pmatrix}$$

$$a^k = (\times \quad\ \ \times \qquad\qquad\qquad) \qquad a^k = (\otimes \ \ \oplus \ \ \otimes \quad\ \ \oplus \qquad\quad)$$

Before rotation *After rotation*

FIG. 1. An example illustrating the annihilation process.

introduced into R^k and a^k due to the annihilation process are denoted by $+$, and all elements involved in the transformation are circled. Of course elements in a^k denoted by $\oplus$ are ultimately annihilated.

The ascending sequence of row indices involved in the *elimination* of a^k is referred to as its *rotation sequence*, and we denote it by $\Xi^k = \{\xi_1^k, \xi_2^k, \ldots, \xi_{\mu_k}^k\}$. (In [8], the rotation sequence is called the *annihilation sequence*.) Here μ_k is the *length* of Ξ^k. In Figure 1, $\Xi^k = \{1,2,3,5\}$ and its length is 4. Clearly $1 \leqslant \xi_j^k \leqslant n$ for $1 \leqslant j \leqslant \mu_k$.

Note that the rotation sequence terminates for one of two reasons:

(1) Row $\xi_{\mu_k}^k$ of R^k has no off-diagonal nonzeros, as in the example of Figure 1.

(2) Row $\xi_{\mu_k}^k$ of R^{k-1} is empty. Hence the row being annihilated can be transferred into row $\xi_{\mu_k}^k$ of R^{k-1}.

In case (1), the rotation sequence is said to be *maximal*. When $m \gg n$, rotation sequences will tend to be maximal. Hence, throughout the discussion in this paper, it is convenient to assume that each rotation sequence is maximal.

Note that the number of rotations required to annihilate row a^k is exactly μ_k. Let M be a matrix. We denote row p of M by M_{p*} and the number of nonzeros in M_{p*} by $|M_{p*}|$. Now, using the definition of the cost of annihilating a nonzero, the *cost* of annihilating row a^k is therefore

$$\sum_{p \in \Xi^k} |R_{p*}^k|. \tag{2.1}$$

That is, the cost depends on the structure of R^k, which depends on the structures of R^{k-1} and a^k. In other words, the cost depends on the structures of the first k rows of A. Note that if the rotation of a row into R^{k-1} simply amounts to transferring it into R^{k-1}, the cost of such an operation, according to (2.1), is also given by the number of nonzeros in the given row. We adopt this definition for simplicity, and because in most cases time proportional to the number of nonzeros in the row will be expended, even if it is not done in performing arithmetic.

The discussion above shows that the cost of annihilating the rows of A depends on the order in which they are processed. It is also well known that the sparsity of the final upper triangular matrix R (or R^m) depends only on the ordering of the columns of A [3]. Thus, for efficient implementation of the annihilating process, it is desirable to find column and row orderings so that R is sparse and the cost of computing it is small.

3. AN IMPLICIT GRAPH MODEL FOR ROW ANNIHILATION AND SOME BASIC RESULTS

Recall from Section 2 that the cost of annihilating a row, say a^k, depends on the structures of a^k and R^{k-1}. (Implicitly the cost depends on the structures of the first k rows of A or the row ordering of A.) It is therefore desirable for a model for studying the row-annihilation process and the row-ordering problem to possess the following characteristics:

(1) It models the structures of the sequence of upper triangular matrices $R^0, R^1, \dots, R^{m-1}, R^m$.

(2) It provides a way to simulate the process of rotating a row, say a^k, into R^{k-1}.

Note that the sequence $\{R^0, R^1, R^2, \dots, R^{m-1}, R^m\}$ effectively "approaches" R. Since

$$A^TA = \begin{pmatrix} R^T & O \end{pmatrix} Q^T Q \begin{pmatrix} R \\ O \end{pmatrix} = R^T R,$$

R is mathematically the Cholesky factor of the symmetric positive definite matrix A^TA (apart from possible sign differences in some rows). Thus a useful model for studying the row-annihilation and row-ordering problems would be one that eventually "converges" to a model for sparse Cholesky decomposition. By way of preparation, we will first review briefly a graph model for sparse Cholesky factorization. In the following discussion, if M is a matrix, M_{ij} denotes the (i, j) element of M, and $|M|$ denotes the number of nonzeros in M. Also $|S|$ denotes the number of elements in S if S is a set.

Let B be an $n \times n$ sparse symmetric positive definite matrix. The *symmetric graph* (or simply *graph*) of B, denoted by $G(B)=(X(B), E(B))$, is a labeled undirected graph with $X(B)=\{x_1, x_2, \ldots, x_n\}$ and $\{x_i, x_j\} \in E(B)$ if and only if $B_{ij} \neq 0$ for $i \neq j$. Here x_i is the node having label i. Sparse Cholesky decomposition can be described using the graph of B [5].

Let S be a subset of $X(B)$. A node $y \in X(B)-S$ is said to be *reachable from* $x \in X(B)-S$ *through* S if there is a path $(x, v_1, v_2, \ldots, v_l, y)$ from x to y in $G(B)$ with $v_i \in S$ for $1 \leqslant i \leqslant l$. Note that S may be empty and l may be zero. The set of nodes that are reachable from x through S, denoted by $\mathrm{Reach}_{G(B)}(x, S)$, is called the *reachable set* of x through S.

Let R be the Cholesky factor of B. Lemma 3.1 tells us exactly where fill-in will occur during the sparse Cholesky decomposition of the matrix B, in terms of its graph [5, 11].

LEMMA 3.1. *Let* $S_i=\{x_1, x_2, \ldots, x_{i-1}\}$. *Then for* $j>i$, $R_{ij} \neq 0$ *if and only if* $x_j \in \mathrm{Reach}_{G(B)}(x_i, S_i)$.

Consequently the number of off-diagonal nonzeros in R is

$$|R|=\sum_{i=1}^{n}\left|\mathrm{Reach}_{G(B)}(x_i, S_i)\right|.$$

Suppose P is any $n \times n$ permutation matrix. Let $\bar{B}=P^T B P$, and denote the symmetric graph of $\bar{B}$ by $G(\bar{B})=(X(\bar{B}), E(\bar{B}))$. Let $X(\bar{B})=\{\bar{x}_1, \bar{x}_2, \ldots, \bar{x}_n\}$, where $\bar{x}_i$ is the node having label i in $G(\bar{B})$. The *structures* of $G(B)$ and $G(\bar{B})$ are identical, but their *labelings* are different.

Let $\bar{R}$ be the Cholesky factor of $\bar{B}$. Then the number of off-diagonal nonzeros in $\bar{R}$ is

$$|\bar{R}|=\sum_{i=1}^{n}\left|\mathrm{Reach}_{G(\bar{B})}(\bar{x}_i, \bar{S}_i)\right|,$$

where $\bar{S}_i=\{\bar{x}_1, \bar{x}_2, \ldots, \bar{x}_{i-1}\}$. Clearly $|R|$ and $|\bar{R}|$ are in general different. That is, $|\bar{R}|$ depends on the labeling of the nodes (or the symmetric ordering of the rows and columns of B). Hence the problem of finding a "good" permutation for a sparse symmetric positive definite matrix B can be stated as follows. Given the graph $G(B)$ of B, relabel the nodes of $G(B)$ so that the reachable sets $\mathrm{Reach}_{G(\bar{B})}(\bar{x}_i, \bar{S}_i)$ are small. A thorough treatment of this problem can be found in [5].

Now we describe an implicit graph model for studying the row-transformation and the row-ordering problems. For each row a^k of A, define an

$n \times n$ symmetric matrix $Y^k = (a^k)^T(a^k)$. Note that $Y_{ij}^k \neq 0$ if and only if a_i^k and a_j^k are both nonzero. Thus, some of the columns and rows of Y^k may be null; they correspond to zeros in a^k. In fact, if *all* the null columns and null rows are deleted from Y^k, what is left is a dense square matrix whose order is the same as the number of nonzeros in a^k.

The *row graph* of a^k, denoted by $\phi^k = (\chi^k, \epsilon^k)$, is a labeled undirected graph with $\chi^k = \{x_i | a_i^k \neq 0\}$ and $\epsilon^k = \{\{x_i, x_j\} | x_i, x_j \in \chi^k\}$. Hence the row graph of any row of A is a complete graph, and it is the symmetric graph of the dense submatrix in Y^k. The row graph ϕ^k of a^k will sometimes be called the *graph* of Y^k. An example is given in Figure 2.

Recall that the cost of annihilating a^k depends on the structures of the first k rows of A, so it is helpful to consider the k-by-n matrix A^k which is

$$A = \begin{pmatrix} \times & \times & \times & & & & \\ & & & \times & & & \times \\ \times & & & & & & \\ & & & & \times & & \times \\ & \times & \times & & & & \\ & & & \times & \times & & \\ \times & & & & \times & \times & \\ \times & & \times & & \times & & \\ & & & & & \times & \times \\ & \times & & \times & & & \end{pmatrix}$$

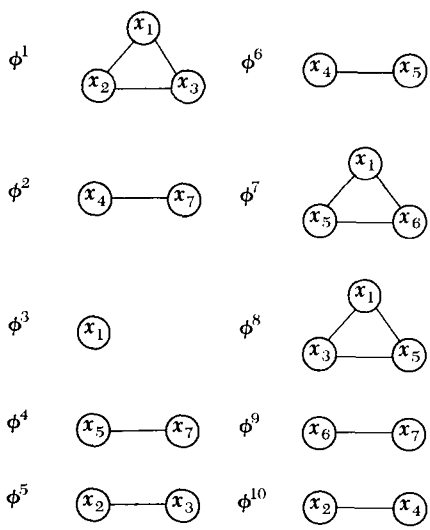

FIG. 2. The sequence of row graphs of a 10-by-7 matrix.

defined by

$$A^k = \begin{pmatrix} a^1 \\ a^2 \\ \vdots \\ a^k \end{pmatrix}.$$

The upper triangular matrix R^k is obtained by annihilating the rows of A^k using rotations. We will assume that exact cancellation does not occur and all rotation sequences are maximal. Let B^k denote the $n \times n$ symmetric matrix $(A^k)^T(A^k)$. Note that

$$B^k = \begin{pmatrix} (a^1)^T & (a^2)^T & \cdots & (a^k)^T \end{pmatrix} \begin{pmatrix} a^1 \\ a^2 \\ \vdots \\ a^k \end{pmatrix} = \sum_{l=1}^{k} (a^l)^T (a^l) = \sum_{l=1}^{k} Y^l.$$

That is, the symmetric matrix B^k is the sum of k sparse matrices, each of which contains a dense submatrix. Lemma 3.2 follows directly from the definition of B^k.

LEMMA 3.2. *Assume exact cancellation does not occur. Then* $B^k_{ij} \neq 0$ *if and only if* $a^l_i \neq 0$ *and* $a^l_j \neq 0$ *for some* $l \leq k$.

Note that there may be some null columns in the matrix A^k. Thus as in the case of Y^l, B^k may be structurally singular, since row l and column l of B^k are null whenever column l of A^k is null.

Define a labeled undirected graph $G^k = (X^k, E^k)$ as follows. Let

$$X^k = \left\{ x_i \,|\, \text{column } i \text{ and row } i \text{ of } B^k \text{ are non-null} \right\},$$

and

$$E^k = \left\{ \{x_i, x_j\} \,|\, B^k_{ij} \neq 0 \right\}.$$

The sequence of G^k corresponding to the example given in Figure 2 is shown in Figure 3. Let $\hat{B}^k$ be the submatrix obtained from B^k by deleting all the null rows and null columns. Then G^k is the symmetric graph of $\hat{B}^k$. If the non-null columns of A^k are linearly independent, then $\hat{B}^k$ is symmetric and positive definite.

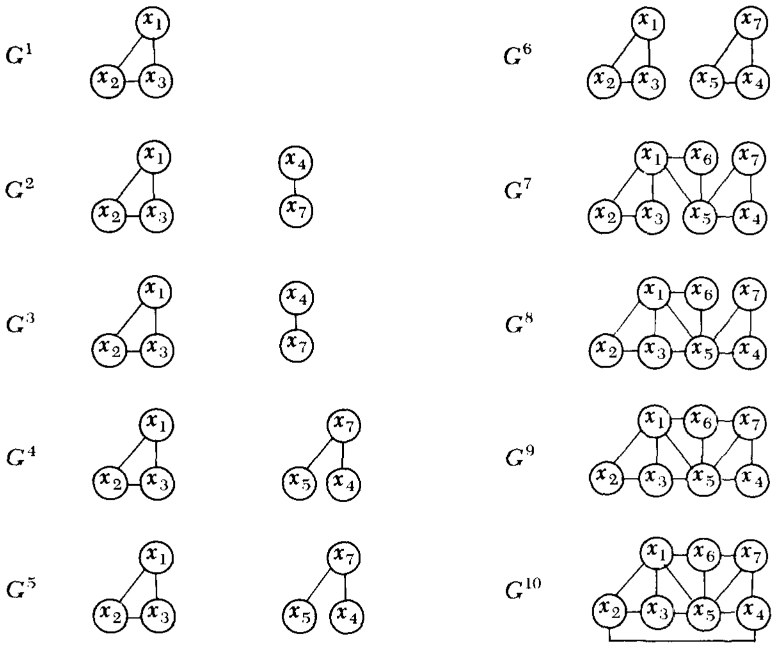

FIG. 3. The sequence of G^k for the matrix shown in Figure 2.

With this framework, it follows that G^k can be defined using the row graphs of a^1, $a^2, \ldots,$ and a^k, which we state as a lemma.

LEMMA 3.3. $G^k = \bigcup_{l=1}^{k} \phi^l$. *That is,* $X^k = \bigcup_{l=1}^{k} \chi^l$ *and* $E^k = \bigcup_{l=1}^{k} \epsilon^l$.

Proof. Suppose $x_i, x_j \in X^k$ and $\{x_i, x_j\} \in E^k$. Then by the definition of G^k, $B_{ij}^k \neq 0$. Applying Lemma 3.2, there must exist some $l \leqslant k$ such that $a_i^l \neq 0$ and $a_j^l \neq 0$. Hence by the definition of row graphs, $x_i, x_j \in \chi^l$ and $\{x_i, x_j\} \in \epsilon^l$.

On the other hand, if $x_i, x_j \in \chi^l$ and $\{x_i, x_j\} \in \epsilon^l$, for $l \leqslant k$, then a_i^l and a_j^l are both nonzero. Lemma 3.2 and the definition of G^k immediately imply that $x_i, x_j \in X^k$ and $\{x_i, x_j\} \in E^k$. ∎

Note that when $k = m$ we have $B^m = (A^m)^T(A^m) = A^TA$, and $G^m = \bigcup_{l=1}^{m} \phi^l$ becomes the graph of A^TA. Thus, assuming exact cancellation does not occur, the nonzero structure of the Cholesky factor R^m (or R) of the symmetric positive definite matrix A^TA can be obtained from G^m using Lemma 3.1. Also note that relabeling the nodes of G^m (i.e., reordering the rows and columns of

B^m) is the same as reordering the columns of A, since

$$P^T B^m P = P^T A^T A P = (AP)^T (AP).$$

What is the structure of R^k when $k < m$? Note that B^k may be structurally singular, since it may have null columns and null rows. But as we have pointed out earlier, any null columns and null rows in B^k must correspond to null columns and null rows in R^k. In fact if $\hat{R}^k$ is the submatrix obtained from R^k by deleting those null columns and null rows, then $\hat{R}^k$ is structurally the Cholesky factor of $\hat{B}^k$, assuming all rotation sequences are maximal. Thus, apart from the null columns and null rows in R^k, the nonzero structure of R^k can be determined from G^k using Lemma 3.1. Here instead of using S_i, we have S_i^k which includes only those nodes that are actually in G^k. This is summarized in the following lemma, which is a generalization of Lemma 3.1.

LEMMA 3.4. *Assume exact cancellation does not occur and all rotation sequences are maximal. Then for $j > i$, $R_{ij}^k \neq 0$ if and only if $x_i, x_j \in X^k$ and $x_j \in \mathrm{Reach}_{G^k}(x_i, S_i^k)$, where $S_i^k = \{x_l \in X^k \mid l < i\}$.*

Thus the structures of the upper triangular matrices $R^1, R^2, \ldots, R^m$ can be determined by applying Lemma 3.4 to the sequence of graphs $G^1, G^2, \ldots, G^m$. It is important to note that our discussion and Lemmas 3.1 and 3.4 assume that we are working with B^k and that exact cancellation does not occur during the computation. Furthermore all rotation sequences are assumed to be maximal. The structure of R^k, $k = 1, 2, \ldots, m$, determined under these assumptions only represents the worst-case situation. This is illustrated by two examples in Figures 4 and 5. If $m \gg n$, there is likely to be more maximal rotation sequences than nonmaximal ones. Thus in this case the actual structure of R should be close to that predicted by Lemma 3.1. Experience shows that this is indeed the case [10]. Furthermore, the model presented here predicts *exactly* the structures of some of the triangular matrices, including R, for some important classes of problems [10].

We now consider the graph G^k in more detail. By Lemma 3.3, we have

$$G^k = \bigcup_{l=1}^{k} \phi^l = \left\{ \bigcup_{l=1}^{k-1} \phi^l \right\} \cup \phi^k = G^{k-1} \cup \phi^k.$$

Thus from the graph-theory point of view, annihilating a row is equivalent to *merging the row graph ϕ^k with the existing graph G^{k-1} of B^{k-1} and applying Lemma* 3.4 *to G^k*. Lemma 3.5 relates the rotation sequence we

$$A^3 = \begin{pmatrix} \times & \times & \times & & \\ & \times & & \times & \\ & & & & \times \end{pmatrix}$$

$$B^3 = \begin{pmatrix} \times & \times & \times & & \\ \times & \times & \times & \times & \\ \times & \times & \times & & \\ & \times & & \times & \\ & & & & \times \end{pmatrix}$$

$$\text{Expected structure of } R^3 = \begin{pmatrix} \times & \times & \times & & \\ & \times & \times & \times & \\ & & \times & \times & \\ & & & \times & \\ & & & & \times \end{pmatrix}$$

$$\text{Actual structure of } R^3 = \begin{pmatrix} \times & \times & \times & & \\ & \times & & \times & \\ & & & & \\ & & & & \\ & & & & \times \end{pmatrix}$$

FIG. 4. Difference between the expected and the actual structures of R^k for $k = 3$.

$$A = \begin{pmatrix} \times & \times & \times & \times & \times \\ & \times & & & \\ & & \times & & \\ & & & \times & \\ & & & & \times \end{pmatrix}, \qquad B = \begin{pmatrix} \times & \times & \times & \times & \times \\ \times & \times & \times & \times & \times \\ \times & \times & \times & \times & \times \\ \times & \times & \times & \times & \times \\ \times & \times & \times & \times & \times \end{pmatrix}$$

$$\text{Expected structure of } R = \begin{pmatrix} \times & \times & \times & \times & \times \\ & \times & \times & \times & \times \\ & & \times & \times & \times \\ & & & \times & \times \\ & & & & \times \end{pmatrix}$$

$$\text{Actual structure of } R = \begin{pmatrix} \times & \times & \times & \times & \times \\ & \times & & & \\ & & \times & & \\ & & & \times & \\ & & & & \times \end{pmatrix}$$

FIG. 5. Another example illustrating the difference between the expected and the actual structures of R.

introduced in Section 2 to our graph model. It also shows that the rotation sequence depends on the structure of G^k and the labeling of X^k. Note that we are assuming that exact cancellation does not occur and all rotation sequences are maximal.

LEMMA 3.5. *Let* $\Xi^k = \{\xi_1^k, \xi_2^k, \ldots, \xi_{\mu_k}^k\}$ *be the rotation sequence of row* a^k. *Then*:

(1) $\xi_1^k = \min\{l \mid x_l \in \chi^k\}$.

(2) *For* $1 \leqslant i < \mu_k$, $\xi_{i+1}^k = \min\{l \mid x_l \in \mathrm{Reach}_{G^k}(x_{\xi_i^k}, S_{\xi_i^k}^k)\}$, *where* $S_{\xi_i^k}^k = \{x_j \in X^k \mid j < \xi_i^k\}$.

Proof. (1): The component in a^k to be annihilated first is the first nonzero in a^k. Since ϕ^k is the row graph of a^k, the node x_p, where $p = \min\{l \mid x_l \in \chi^k\}$, corresponds to the first nonzero in a^k.

(2): Consider ξ_l^k, $1 \leqslant l \leqslant \mu_k$. Clearly ξ_{l+1}^k is the column index of the first off-diagonal nonzero in row ξ_l^k of R^k. By Lemma 3.4, the off-diagonal nonzeros in row ξ_i^k of R^k are given by the set $\mathrm{Reach}_{G^k}(x_{\xi_i^k}, S_{\xi_i^k}^k)$. Hence the result follows. ∎

In Section 2 the cost of annihilating a^k is defined to be $\sum_{p \in \Xi^k} |R_{p*}^k|$. This can be restated using our graph model: the cost of annihilating a^k is

$$\theta^k = \sum_{p \in \Xi^k} \left\{ \left| \mathrm{Reach}_{G^k}\left(x_p, S_p^k\right) \right| + 1 \right\}.$$

Note that this cost depends on the structure of G^k, which depends, in turn, on the structures of the row graphs $\phi^1, \phi^2, \ldots, \phi^k$. Thus the problem of finding a "good" row ordering for A can now be viewed as the problem of ordering the row graphs so that the quantities θ^k, $1 \leqslant k \leqslant m$, are small.

Some basic results about the row-annihilation (or graph-merging) process can be derived easily from this graph model.

Let $\psi^k = \{C_1^k, C_2^k, \ldots, C_{\gamma_k}^k\}$ be the *component partitioning* of G^k; that is, the partitioning of the node set X^k induced by the connected components $G^k(C_j^k)$ of the graph G^k. Thus for $1 \leqslant i$, $j \leqslant \gamma_k$ and $i \neq j$, $C_i^k \cap C_j^k = \varnothing$ and $\bigcup_{l=1}^{\gamma_k} C_l^k = X^k$. Furthermore there is no path joining nodes in C_i^k and nodes in C_j^k when $i \neq j$.

It is important to note that the sequence of component partitionings $\{\psi^k\}$ depends *only* on the order in which the row graphs are merged; that is, it depends on the *row ordering* of A. The effect of permuting the columns of A is just a relabeling of the nodes of G^k, which may, however, affect the cost of annihilating the rows, and the nonzero structures of the upper triangular matrices.

The next result illustrates the significance of the component partitioning ψ^k. It is also stated in [8] using matrix notation, but the proof given here is based on the implicit model and is simpler than the one given in [8].

THEOREM 3.6. *Let $\psi^k = \{C_1^k, C_2^k, \dots, C_{\gamma_k}^k\}$ be the component partitioning of G^k. Denote the component containing the row graph ϕ^k by $G^k(C_{\sigma_k}^k)$. Then $x_p \in C_{\sigma_k}^k$ for all $p \in \Xi^k$.*

Proof. The proof is by contradiction. Since $x_{\xi_1^k} \in C_{\sigma_k}^k$, there must exist two consecutive members s and t of Ξ^k such that $x_s \in C_{\sigma_k}^k$ and $x_t \in C_l^k$, for some $l \neq \sigma_k$. From Lemma 3.4, $t = \min\{i \mid x_i \in \text{Reach}_{G^k}(x_s, S_s^k)\}$. This means that there must be a path joining x_s and x_t in G^k. This contradicts the fact that x_s and x_t are in different components. ■

Theorem 3.6 shows that the set of nodes involved in the annihilation of row k (that is, the nodes corresponding to the pivot elements),

$$\{x_{\xi_1^k}, x_{\xi_2^k}, \dots, x_{\xi_{\mu_k}^k}\},$$

is limited to the set $C_{\sigma_k}^k$ which contains the node set χ^k of the row graph ϕ^k. The next result follows from the observation that the cost of annihilating a row depends in part on the *length* of the rotation sequence.

LEMMA 3.7. *The cost of annihilating row k is bounded by*

$$\tfrac{1}{2}\mu_k(\mu_k + 1) \leqslant \tfrac{1}{2}|C_{\sigma_k}^k|\left(|C_{\sigma_k}^k| + 1\right).$$

Hence we want to find a row ordering which allows the component $G^k(C_{\sigma_k}^k)$ to be kept small for as large a k as possible. Corollary 3.8 provides a tighter bound on the cost of the annihilation.

COROLLARY 3.8. *Let Ξ^k be the rotation sequence of row a^k, and $\eta = \min\{p \mid p \in \Xi^k\}$. Then $\Xi^k \subseteq \{q \geqslant \eta \mid x_q \in C_{\sigma_k}^k\}$. Furthermore, if $\delta_k = |\{q \geqslant \eta \mid x_q \in C_{\sigma_k}^k\}|$, then the cost of annihilating a^k is bounded by $\frac{1}{2}\delta_k(\delta_k + 1)$.*

The result above says that regardless of whether $C_{\sigma_k}^k$ is small, we should arrange for the column index of the first nonzero of row k to be as large as possible.

4. ROW-ORDERING STRATEGIES FOR NESTED-DISSECTION COLUMN ORDERINGS

The results in the previous section provide us with some insight into the problem of ordering the rows. Theorem 3.6 says that the rows should be arranged so that the node sets $\{C^1_{\sigma_1}, C^2_{\sigma_2}, \ldots, C^m_{\sigma_m}\}$ can be kept small. Of course, Lemma 3.4 also says that the columns should be ordered so that the amount of fill-in is small. In this section we use the implicit graph model and the results to show that good row orderings can be obtained from both the conventional (width-1) and width-2 nested-dissection column orderings.

In the following discussion, we denote the symmetric graph of A^TA by $G=(X,E)$. For simplicity we assume that G is connected. Let $S \subseteq X$. Suppose the nodes of S, together with the incident edges, are removed from G. If the remaining graph is disconnected, then S is called a *separator*. Let C_1 and C_2 denote the node sets of two distinct connected components in the remaining graph. If the distance between any x in C_1 and any y in C_2 in the original graph G is greater than l, then S is a *width-l separator*. An ordering of X is a *width-l dissection ordering* if x is labeled after y for $x \in S$ and $y \in X - S$. This labeling technique can be applied recursively to order the nodes of $X - S$, resulting in the *width-l nested-dissection ordering*. In this paper, we are interested in the cases where $l = 1$ and $l = 2$.

Width-1 nested-dissection orderings are discussed in detail in [2,4] in connection with the solution of sparse symmetric positive definite systems, and width-2 nested dissection orderings are considered in [8] for finding good row and column orderings for solving sparse linear least-squares problems. Gilbert also employs width-2 separators (or wide separators) in the solution of sparse systems of linear equations using Gaussian elimination with partial pivoting [9].

In the following discussion, the column index of the first (last) nonzero in a row is called the *first* (*last*) *subscript* of that row.

4.1. *Width-2 Nested Dissection*

In [8], George and Ng have shown that if the labeling of X is a width-2 nested-dissection labeling, then a good row ordering can be obtained simply by arranging the rows of the matrix so that the first subscripts are in ascending order. We refer to this row ordering as the row ordering *induced* by a width-2 nested-dissection column ordering. Denote the permuted matrix by A. The key to the success of this row-ordering strategy is in the following result [8].

LEMMA 4.1. *Let S be a minimal width-2 separator of G. Then S is the union of one or more row-graph vertex sets.*

Thus, implicitly, a minimal width-2 separator identifies a set of rows Z_S. More precisely, Z_S contains those rows a^k of A such that the vertex sets χ^k of the associated row graphs ϕ^k are contained in S. The next result follows directly from the definition of width-2 separators and the way in which the graph G is formed.

COROLLARY 4.2. *If the row graphs associated with the rows of* Z_S *are removed from* G, *the remaining graph is disconnected.*

We now look at the strategy more carefully. Assume S is the first minimal width-2 separator chosen from G in width-2 nested dissection. For simplicity, assume the graph remaining after the row graphs associated with the rows of Z_S have been removed has two components. Consider two rows a^i and a^j, and denote their row graphs respectively by $\phi^i = (\chi^i, \epsilon^i)$ and $\phi^j = (\chi^j, \epsilon^j)$. Suppose $\chi^i \subseteq S$ and $\chi^j \nsubseteq S$. (Note that $\chi^i \cap \chi^j$ and $\chi^j \cap S$ are not necessarily empty.) That is, $a^i \in Z_S$ and $a^j \notin Z_S$. Let $x_p \in \chi^i$ and $x_q \in \chi^j$, where

$$p = \min\{l \mid x_l \in \chi^i\} \quad \text{and} \quad q = \min\{l \mid x_l \in \chi^j\}.$$

That is, x_p and x_q are vertices whose labelings are the first subscripts in the corresponding rows. Since the vertex ordering is a width-2 nested-dissection ordering, $x_p \in S$, $x_q \notin S$, and $p > q$. Thus, row a^i will be annihilated *after* row a^j. In fact, the row-ordering strategy implies that the rows of Z_S will be annihilated *last*.

Now consider the annihilation of row a^j ($a^j \notin Z_S$). This row will be annihilated before the rows of Z_S. Let $G^j = (X^j, E^j)$ be the union of the row graphs of those rows that have been annihilated. In the worst case, G^j is the graph obtained by removing the row graphs associated with the rows of Z_S; this corresponds to the case when only the rows of Z_S have not been annihilated. By Corollary 4.2, G^j is disconnected in this worst case. In general, G^j is a proper subgraph of G. This is illustrated in Figure 6.

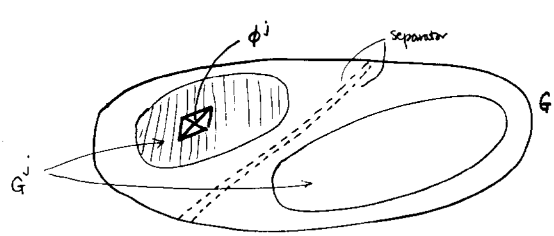

FIG. 6. Width-2 nested dissection: annihilation of a row $a^j \notin Z_S$. The component containing ϕ^j is shaded.

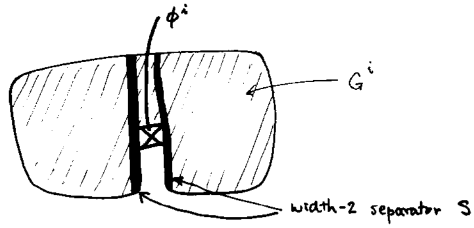

FIG. 7. Width-2 nested dissection: annihilation of a row $a^i \in Z_S$. The component containing ϕ^i is shaded.

Applying Theorem 3.6, the component containing χ^j is therefore a *subgraph* of G^j. Hence the length of the rotation sequence of a^j and (consequently by Lemma 3.7) the cost of annihilating a^j are both limited.

Consider the annihilation of row a^i ($a^i \in Z_S$). Let $G^i = (X^i, E^i)$ be the union of the row graphs of those rows that have been annihilated. At this point, all the rows of A, except some of those of Z_S, have been annihilated. By the definition of S, the graph G^i is connected (see Figure 7), and hence, by Theorem 3.6, the length of the rotation sequence could be as large as the number of vertices in X^i, which could be the same as X. In view of Lemma 3.7, this is certainly undesirable. However, since the vertex labeling is a nested-dissection labeling, from the discussion above it follows that the node corresponding to the first subscript of a^i must be in S. Therefore, following from Lemma 3.4, the rotation sequence is limited to the vertices of S. Thus, applying Corollary 3.8, the cost of annihilating a^i cannot be greater than $\frac{1}{2}|S|(|S|+1)$.

Because of the recursive nature of the column ordering, one can see that the same argument can be applied repeatedly to rows in the components obtained when the row graphs associated with the rows of Z_S are removed. In general, if χ^i is a subset of a width-2 separator obtained on the kth level of recursion, and if the induced row ordering is used, then only the vertices of this width-2 separator will be involved in the annihilation of row a^i. The cost of annihilating a^i is therefore small. Thus the row ordering induced by a width-2 nested dissection column ordering is indeed a good row ordering.

4.2. *Width-1 Nested Dissection*

We now use the results derived in Section 3 to show that a good row ordering can also be obtained easily from a width-1 nested-dissection column ordering. Let S be the first minimal width-1 separator obtained in width-1 nested dissection. For simplicity, assume the graph obtained when the nodes of S and the incident edges are removed has exactly two connected components. Denote the vertex sets of the components by C_1 and C_2.

LEMMA 4.3 [8]. *Let χ be a clique in G. Then either $\chi \subseteq S \cup C_1$ or $\chi \subseteq S \cup C_2$.*

The next result is important, since it says that a width-1 separator also *identifies* a set of rows implicitly.

LEMMA 4.4. $S \cup \mathrm{Adj}(S)$ *is the union of one or more row-graph vertex sets.*

Proof. Let $x \in S$. Because of the way G is defined, there is at least one row graph, say $\phi^k = (\chi^k, \epsilon^k)$, such that $x \in \chi^k$. The result follows, since χ^k is a clique and $\chi^k \subseteq \mathrm{Adj}(x) \cup \{x\} \subseteq S \cup \mathrm{Adj}(S)$. ∎

Thus, a width-1 separator, *together with its adjacent set*, identifies a set of rows. Denote this set of rows by Z_S. More specifically, Z_S contains those rows a^k such that each of the vertex sets χ^k of the associated row graphs ϕ^k has a nonempty intersection with S. The following corollary follows from the definition of S and the way in which G is formed.

COROLLARY 4.5. *If the row graphs associated with the rows of Z_S are removed from G, the remaining graph is disconnected.*

Now consider two rows, say rows a^i and a^j, whose row graphs are respectively $\phi^i = (\chi^i, \epsilon^i)$ and $\phi^j = (\chi^j, \epsilon^j)$. Suppose $\chi^i \cap S \neq \varnothing$ and $\chi^j \cap S = \varnothing$. (Again, $\chi^i \cap \chi^j$ may not be empty.) That is, $a^i \in Z_S$ and $a^j \notin Z_S$.

First consider the annihilation of row a^j. Suppose we annihilate a^j before any rows of Z_S. Let $G^j = (X^j, E^j)$ be the union of the row graphs of those rows that have been annihilated. The graph G^j is, in the worst case, the graph that would be obtained when the row graphs associated with the rows of Z_S are removed from G. Even in this worst case, G^j would be disconnected. This is illustrated in Figure 8. From the discussion above, G^j is a proper subgraph of G. Applying Theorem 3.6, the component containing χ^j is a subgraph of G^j and hence a proper subgraph of G. The length of the rotation

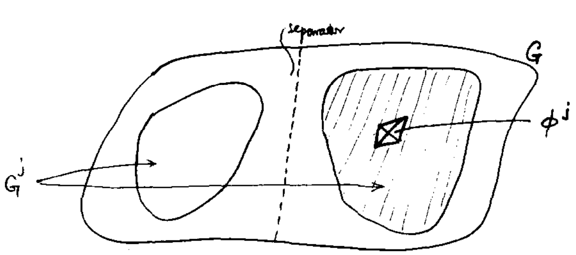

FIG. 8. Width-1 nested dissection: annihilation of a row $a^j \notin Z_S$. The component containing ϕ^j is shaded.

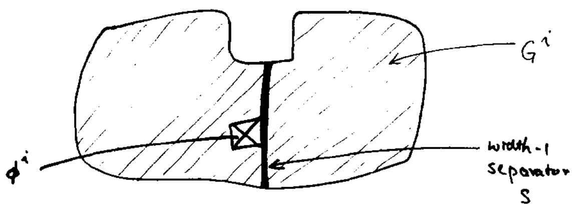

FIG. 9. Width-1 nested dissection: annihilation of a row $a^i \in Z_S$. The component containing ϕ^i is shaded.

sequence of a^j, and consequently (by Lemma 3.7) the cost of annihilating a^j, are both limited.

Consider the annihilation of row a^i. We assume that a^i will be annihilated after all the rows not in Z_S have been annihilated. Let $G^i = (X^i, E^i)$ be the union of the row graphs of those rows that have been annihilated. From Lemma 4.3, χ^i is in $S \cup C_1$ or $S \cup C_2$. The vertex set of the component that contains χ^i may be as large as X^i. This is illustrated in Figure 9. Moreover, suppose $x_t \in \chi^i$, where $t = \min\{l \mid x_l \in \chi^i\}$. (That is, x_t is the node corresponding to the first subscript of row a^i.) There are three possibilities: $x_t \in S$, $x_t \in C_1$, and $x_t \in C_2$. In the first case, we see that $\chi^i \subseteq S$, since the vertex labeling is a width-1 nested-dissection labeling and S is a width-1 separator. Thus, according to Lemma 3.4, the rotation sequence cannot involve vertices other than those of S. Hence, applying Corollary 3.8, the cost of annihilating a^i cannot be larger than $\frac{1}{2}|S|(|S|+1)$.

Now suppose $x_t \in C_1$ or $x_t \in C_2$. Since the vertex labeling is a width-1 nested-dissection ordering and S is a separator, t is smaller than the labeling of any node of S. In fact, it is possible that t is *much* smaller than any labeling of the nodes of S. Applying Corollary 3.8, the cost of annihilating a^i may be large, since there may be many vertices in C_1 and/or C_2 whose labelings are greater than t. Thus, annihilating the rows of Z_S last does not *appear* to be an effective strategy.

However, it *is* an effective strategy, as the following discussion shows. Recall that the vertices of C_1 and C_2 are labeled using the same strategy recursively. We will call a separator obtained on the kth level of recursion a level-k separator. An example is shown in Figure 10, in which a level-k separator is denoted by $\langle k \rangle$. Suppose x_t belongs to a level-p separator, where $p > 1$. Applying Lemma 3.4, we see that in the worst case, all the vertices in this level-p separator are reachable from x_t. Assuming the annihilation sequence is maximal, these vertices will also be part of the rotation sequence. See Figure 11 for an example. Also note that the level-p separator is obtained by applying the dissection technique to a vertex set $\bar{C}$ of a component on the

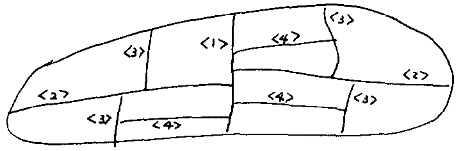

FIG. 10. Width-1 nested dissection. Level-k separators are indicated by $\langle k \rangle$.

$(p-1)$st level of recursion, and this component is surrounded by portions of level-q separators, with $q < p$, one of which is a level-$(p-1)$ separator. In Figure 11, x_t belongs to a level-6 separator, and the vertex set $\bar{C}$ is surrounded by portions of level-q separators, where $q = 1$, 2, 4, and 5. Using Lemma 3.4 and recalling the fact that the vertices of a level-q separator have larger labels than those of a level-p separator, the reachable set of any vertex on the level-p separator must include the vertices in the portions of the level-q separators mentioned above, and these vertices will also be part of the rotation sequence. No vertices on other level-p or level-s $(s > p)$ separators are reachable from x_t. This is a consequence of the (vertex) labeling strategy. In order to generate the entire rotation sequence, we next determine the reachable set of each of the vertices in the portion of the level-$(p-1)$ separator surrounding $\bar{C}$. In Figure 11, we have to examine the reachable sets of the vertices on the level-5 separator marked $\langle 5 \rangle$. By applying this argument repeatedly to the vertices on the level-$(p-1)$ and subsequent related level-s separators $(s < p)$, one can see that the vertices involved in the annihilation of row a^i include, in the worst case, all the vertices on exactly one level-k separator, $1 \leqslant k \leqslant \beta$, where β is the number of recursions. Furthermore, using Lemmas 4.3 and 4.4, one can show that all these separators will be in either $C_1 \cup S$ or $C_2 \cup S$, depending on whether $x_t \in C_1$ or $x_t \in C_2$. This is illustrated in Figure 11. Thus, if each level-k separator is bounded in size, both the length of the rotation sequence of a^k and the cost of annihilating a^k will be limited. Hence a good row-ordering strategy for a

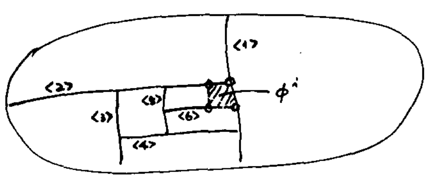

FIG. 11. The vertices involved in the annihilation of a row of Z_S in width-1 nested dissection.

width-1 nested-dissection column ordering is to arrange the rows so that those whose row-graph vertex sets χ^i have nonempty intersection with the separator S are annihilated last.

This idea can be applied recursively to order the rows after a width-1 nested-dissection column ordering is obtained. Denote a separator obtained on the pth level of recursion by S_p. Let $\phi^i = (\chi^i, \epsilon^i)$ be the row graph associated with a^i. Suppose $l = \min\{p \mid \chi^i \cap S_p \neq \varnothing\}$. Then the same argument will show that the rotation sequence will involve all the vertices on exactly one level-k separator, $l \leqslant k \leqslant \beta$.

Thus, if the ordering of the vertices is a width-1 nested-dissection ordering, there is a good row ordering, and as the following discussion indicates, there is a simple way to characterize this ordering. Consider two rows a^i and a^j, and denote their row graphs by $\phi^i = (\chi^i, \epsilon^i)$ and $\phi^j = (\chi^j, \epsilon^j)$. Suppose $\chi^i \cap S \neq \varnothing$ and $\chi^j \cap S = \varnothing$. The goal is to arrange the rows so that row a^j is annihilated before row a^i. Let $x_p \in \chi^i$ and $x_q \in \chi^j$, where

$$p = \max\{l \mid x_l \in \chi^i\} \quad \text{and} \quad q = \max\{l \mid x_l \in \chi^j\}.$$

That is, x_p and x_q are respectively vertices whose labelings are the last subscripts in the corresponding rows. Note that $x_p \in S$, $x_q \notin S$, and $p > q$. That is, the last subscript of row a^i must be larger than the last subscript of row a^j. In fact, the last subscripts of all the rows of Z_S are larger than those of the rows that are not in Z_S. This observation, which is also true for subsequent levels of recursions, can be used to rearrange the rows that are not in Z_S. In other words, a good row ordering can be obtained simply by arranging the rows of the matrix so that the *last* subscripts are in ascending order. We refer to this row ordering as the row ordering *induced* by a width-1 nested-dissection column ordering.

5. CONCLUDING REMARKS

In this paper, we have presented another graph model to study the row-annihilation and row-ordering problems in the QR decomposition of a sparse matrix using Givens rotations. The graph-theoretic results obtained are used to show that good row orderings can be obtained from both width-1 and width-2 nested-dissection column orderings.

The new implicit graph model is different from the bipartite-graph model we introduced in [6], in the sense that rows are not explicitly represented in the implicit graph model, while they are represented in the bipartite-graph model by some nodes. The implicit graph model provides an alternate way of studying the row-annihilation and row-ordering problems. In [7], we will use both models to analyze the width-1 nested-dissection orderings for a model problem.

REFERENCES

1 W. M. Gentleman, Row elimination for solving sparse linear systems and least squares problems, in *Proceedings of the 1975 Dundee Conference on Numerical Analysis* (G. A. Watson, Ed.), Lecture Notes in Mathematics 506, Springer,1975, pp. 122–133.

2 J. A. George, Nested dissection of a regular finite element mesh, *SIAM J. Numer. Anal.* 10:345–363 (1973).

3 J. A. George and M. T. Heath, Solution of sparse linear least squares problems using Givens rotations, *Linear Algebra Appl.* 34:69–83 (1980).

4 J. A. George and J. W. H. Liu, An automatic nested dissection algorithm for irregular finite element problems, *SIAM J. Numer. Anal.* 15:1053–1069 (1978).

5 J. A. George and J. W. H. Liu, *Computer Solution of Large Sparse Positive Definite Systems*, Prentice-Hall, Englewood Cliffs, N.J., 1981.

6 J. A. George, J. W. H. Liu, and E. G. Y. Ng, Row ordering schemes for sparse Givens transformations. I. Bipartite graph model, *Linear Algebra Appl.* 61:55–81 (1984).

7 J. A. George, J. W. H. Liu, and E. G. Y. Ng, Row ordering schemes for sparse Givens transformations. III. Analysis for a model problem, *Linear Algebra Appl.* 75:225–240 (1986).

8 J. A. George and E. G. Y. Ng, On row and column orderings for sparse least squares problems, *SIAM J. Numer. Anal.* 20:326–344 (1983).

9 J. R. Gilbert, Graph separator theorems and sparse Gaussian elimination, Report No. STAN-CS-80-833, Dept. of Computer Science, Stanford Univ., 1980.

10 E. G. Y. Ng, Row elimination in sparse matrices using rotations, Doctoral Dissertation, Research Report CS-83-01, Dept. of Computer Science, Univ. of Waterloo, 1983.

11 D. J. Rose, R. E. Tarjan, and G. S. Lueker, Algorithmic aspects of vertex elimination on graphs, *SIAM J. Comput.* 5:266–283 (1976).

Received 23 September 1983; revised 15 February 1985